APPENDICES

Appendix A. Notation

Table A1. Notation

Notation	Description
N	Total number of entities
i	Entity index; $i = 1, \dots, N$
T	Total number of time points
t	Time index; $t = 1, \dots, T$
$x_i(t)$	Observed value for entity i at time t
$M_i(t)$	Missingness indicator for entity i at time t: 1 if $x_i(t)$ is missing; 0 otherwise
Q	Total number of candidate periodicities (for missingness mechanism)
q	Index of candidate periodicities (for missingness mechanism); $q = 1, \dots, Q$ (ordered by strength); $q = 0$ indicates no periodicity
P_q	q^{th} most significant observed periodicity (for missingness mechanism), $q = 1, 2, \dots, Q$
$\widetilde{M_i}(t)$	Held-out mask at time t for entity i
Ω_i	Set of hold-out timestamps
$\widehat{p_i}(t; \sigma)$	Predicted missingness probability at time t for entity i
$\mathrm{AIC}(q; \boldsymbol{\sigma})$	Akaike Information Criterion of a PMSI model that uses q periodicities
$\boldsymbol{\sigma}$	Kernel hyperparameters $\{\sigma_0, \dots, \sigma_q\}$ for q periodicities (for missingness mechanism)
$\boldsymbol{\sigma}^*_{\boldsymbol{q+1}}$	Optimal imputation parameters of a PMSI model that uses q periodicities (for missingness mechanism)
q^*	Optimal dimensionality of a PMSI model that uses q periodicities (for missingness mechanism)
Q'	Total number of candidate periodicities (for observed data)
q'	Index of candidate periodicities (for observed data); $q' = 1, \dots, Q'$ (ordered by strength); $q' = 0$ indicates no periodicity
$m_i(t')$	Artificial mask
Ω_i'	Set of artificially masked positions for entity i: $\{t' \mid m_i(t') = 1\}$
$\boldsymbol{\sigma}'$	Kernel hyperparameters $\{\sigma_0', \dots, \sigma_{q'}'\}$ for q' periodicities (for observed data)

$x_{i,\mathrm{imp}}(t';\boldsymbol{\sigma}')$	Imputed value returned by PMSI for entity i at original time index t'
$\boldsymbol{\sigma}'^{*}_{\boldsymbol{q'+1}}$	Optimal imputation parameters of a PMSI model that uses q' periodicities (for observed data)
q'^{*}	Optimal dimensionality of a PMSI model that uses q' periodicities (for observed data)

Appendix B. Pseudocode

Algorithm 1: Periodic Multi-Scale Imputation (PMSI)

Inputs:
- N univariate series $\{x_i(t)\}_{i=1,t=1}^{N,T}$
- Missingness indicators $M_i(t) \in \{0,1\}$, where $M_i(t) = 1$ iff $x_i(t)$ is missing

Outputs:
- Imputed series $\{\hat{x}_i(t)\}_{i=1,t=1}^{N,T}$

For missingness mechanism.

Identify top Q periodicities $\{P_1, \ldots, P_Q\}$ of $\{M_i(t)\}$.

for $q \leftarrow 0$ **to** Q **do**

  Fold $M_i(t)$ into a $(q+1)$-D tensor along $\{P_1, \ldots, P_q\}$ for all i.

  Randomly select hold-out timestamps Ω_i and denote held-out masks by $\tilde{M}_i(t)$ for $t \in \Omega_i$.

  Estimate missingness probabilities $\hat{p}_i(t; \boldsymbol{\sigma})$ using separable Gaussian kernels, with $\boldsymbol{\sigma} \in \mathbb{R}_+^{q+1}$.

$$\mathrm{AIC}(q; \boldsymbol{\sigma}) = 2(q+1) - 2\sum_{i=1}^{N} \sum_{t \in \Omega_i} \Big[\tilde{M}_i(t) \ln \hat{p}_i(t; \boldsymbol{\sigma}) + (1 - \tilde{M}_i(t)) \ln\big(1 - \hat{p}_i(t; \boldsymbol{\sigma})\big)\Big].$$

$$\boldsymbol{\sigma}_{q+1}^* = \arg\min_{\boldsymbol{\sigma}} \mathrm{AIC}(q; \boldsymbol{\sigma}), \qquad \mathrm{AIC}^*(q) = \mathrm{AIC}\big(q; \boldsymbol{\sigma}_{q+1}^*\big).$$

end

$$q^* = \arg\min_{q} \mathrm{AIC}^*(q), \qquad \boldsymbol{\sigma}^* = \boldsymbol{\sigma}_{q^*+1}^*.$$

For observed values.

For observed timestamps only, sample synthetic masks $m_i(t) \sim \mathrm{Bernoulli}\big(\hat{p}_i(t; \boldsymbol{\sigma}^*)\big)$ for t with $M_i(t) = 0$.

Create $x_{i,\mathrm{mask}}(t) = x_i(t)$ and hide values where $m_i(t) = 1$ (treat them as missing).

Identify top Q' periodicities $\{P'_1, \ldots, P'_{Q'}\}$ of $\{x_i(t)\}$.

for $q' \leftarrow 0$ **to** Q' **do**

  Fold $x_{i,\mathrm{mask}}(t)$ into $(q'+1)$-D along $\{P'_1, \ldots, P'_{q'}\}$.

  Tune $\boldsymbol{\sigma}' \in \mathbb{R}_+^{q'+1}$ by minimizing $\mathrm{AIC}_{\mathrm{imp}}(q'; \boldsymbol{\sigma}')$ under a Gaussian error model on synthetically hidden entries.

  Record $\boldsymbol{\sigma}'^*_{q'+1}$ and $\mathrm{AIC}^*_{\mathrm{imp}}(q') = \mathrm{AIC}_{\mathrm{imp}}\big(q'; \boldsymbol{\sigma}'^*_{q'+1}\big)$.

end

$$q'^* = \arg\min_{q'} \mathrm{AIC}^*_{\mathrm{imp}}(q'), \qquad \boldsymbol{\sigma}'^* = \boldsymbol{\sigma}'^*_{q'^*+1}.$$

foreach $i = 1, \ldots, N$ **do**

  Fold $x_i(t)$ into $(q'^* + 1)$-D along $\{P'_1, \ldots, P'_{q'^*}\}$ and impute missing entries using $\boldsymbol{\sigma}'^*$.

end

return $\{\hat{x}_i(t)\}_{i=1,t=1}^{N,T}$.

Appendix C. Data Preprocessing and Inclusion Criteria

C1. Data Preprocessing

Averaging Intervals

The original heart rate data, collected at 15-second intervals, was initially averaged into 3-minute intervals using the following formula:

$$\overline{HR}_{3-\min}(u) = \frac{1}{U}\sum_{v=0}^{U-1} HR(t_0 + (uU + v)\Delta t) \tag{C1}$$

where $\Delta t = 15$ seconds, $U = 3$ minutes / 15 seconds $= 12$, t_0 is the reference start time, and u indexes the 3-minute intervals.

Linear Interpolation

To address small gaps of missing data, linear interpolation was applied, limited to intervals up to half an hour. Gaps exceeding this limit were left as missing data. Suppose that valid observations $(t_1, HR(t_1))$ and $(t_2, HR(t_2))$ surround a gap. The interpolated value for the missing times $t(t_1 < t < t_2)$ is computed as:

$$HR_{\text{interp}}(t) = \begin{cases} HR(t_1) + \dfrac{HR(t_2) - HR(t_1)}{t_2 - t_1}(t - t_1), & (t_2 - t_1) \leq 30 \text{ minutes} \\ \text{missing}, & \text{otherwise} \end{cases} \tag{C2}$$

Savitzky-Golay Smoothing

A Savitzky-Golay filter was applied to smooth the data and reduce noise, using a filter window length of 7 and a polynomial order of 2 ($2m + 1 = 7$):

$$\hat{y}_i = \sum_{k=-m}^{m} c_k y_{i+k} \tag{C3}$$

where c_k is a local least-squares fit of a polynomial of degree 2 over the 7 points centered at i. To be specific, for each window from y_{i-3} to y_{i+3}, the coefficients a_0, a_1, a_2 are estimated by minimizing the squared error

$$\min_{a_0, a_1, a_2} \sum_{k=-3}^{3} [y_{i+k} - (a_0 + a_1 k + a_2 k^2)]^2 \tag{C4}$$

Hourly Aggregation

The smoothed heart rate data was further aggregated into hourly intervals:

$$\overline{HR}_{\text{hour}}(h) = \frac{1}{|I_h|}\sum_{i \in I_h} HR_{\text{smooth}}(i) \tag{C5}$$

$HR_{\text{smooth}}(i)$ is the Savitzky-Golay smoothed heart rate at index i, and $|I_h|$ is the total number of data points in hour h.

C2. Inclusion Criteria

Participants were included in the high-adherence cohort if they met the following conditions: (i) no more than 5 of the 28 days had under 16 hours of recorded data; (ii) at least 15 (of the 28 days) included recorded data between midnight and 4 a.m.; and (iii) the longest interval of consecutive missing data did not exceed 6 hours.

Participants were included in the low-adherence cohort if they did not have consecutive missing data over 24 hours (i.e., no data gaps exceeding 24 hours).

Appendix D. Benchmark Imputation Methods

D1. Last Observation Carried Forward (LOCF)

For the last observation carried forward (LOCF) method, the imputation is defined by the simple rule that each missing value at time t is replaced with the last observed value before t. Mathematically, if x_t is missing and x_{t^*} is the most recent non-missing observation with $t^* < t$, then the imputed value is given by $\hat{x}_t = x_{t^*}$.

In our implementation, the LOCF method was applied by first using a forward fill (propagating the last valid observation forward) and then a backward fill to handle any missing values at the beginning of the series. This method implicitly assumes that the underlying process does not change drastically between consecutive time points.

D2. Piecewise Linear Interpolation

For piecewise linear interpolation, the goal is to impute each missing value using a locally fitted linear model. For a missing data point at index i, one first selects a set of observed points within a window of indices (e.g., from $\max(0, i - w)$ to $\min(n, i + w + 1)$). The window size w is a parameter that determines how many neighboring points to consider for the regression. If there are at least two observed points within this window, a simple linear regression is performed by finding parameters a (slope) and b (intercept) that minimize:

$$\min_{a,b} \sum_{j \in O(i)} \left(x_j - (aj + b)\right)^2 \tag{D1}$$

where $O(i)$ denotes the set of indices in the window where the data are observed. The missing value is then imputed as $\hat{x}_i = ai + b$.

In our experiment, we set the window size as 7. If fewer than two observed points are found, this method eventually fills the remaining gaps using forward fill and backward fill.

D3. AutoRegressive Integrated Moving Average (ARIMA)

For the ARIMA method, the underlying model is described by the equation:

$$\Phi(B)(1 - B)^d x_t = \Theta(B)\epsilon_t \tag{D2}$$

where B is the backshift operator, $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$ is the autoregressive part, $(1 - B)^d$ is the differencing operator to achieve stationarity, and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ is the moving average part.

The ARIMA model is thus denoted as ARIMA(p, d, q), where the orders p, d, and q are tuned to best capture the temporal dependencies.

In our approach, the initial series was first imputed using a simple forward and backward fill, and then the differencing order (d) was automatically determined through stationarity tests (i.e., Kwiatkowski–Phillips–Schmidt–Shin), and the optimal autoregressive (AR) and moving-average (MA) orders (p, q) were selected iteratively by minimizing the Akaike information criterion. After fitting, the in-sample predictions from the model were used to replace only the missing entries in the original series.

D4. k-Nearest Neighbors (KNN)

In a two-dimensional context, k-nearest neighbors imputes a missing value by looking at the k nearest neighbors based on a chosen distance metric. The imputed value $\hat{x}_{ij}$ for a missing entry is given by a weighted average:

$$\hat{x}_{ij} = \frac{\sum_{l=1}^{k} w_l x_{ij}^{(l)}}{\sum_{l=1}^{k} w_l} \tag{D3}$$

where $x_{ij}^{(l)}$ are the values of the k nearest neighbors and w_l is a weight inversely related to the distance.

In our experiment, k-nearest neighbors imputation was performed in two dimensions (hourly and daily). The optimal value of the hyperparameter k (number of neighbors) was determined by the lowest bootstrapped RMSE calculations.

D5. Convolutional Neural Networks (CNN)

The CNN-based imputation treats the time series as a two-dimensional tensor (hourly and daily). Specifically, our autoencoder f_θ is a small fully convolutional network with three 3×3 layers to learn both local and periodic patterns; the first convolution maps the two-dimensional input to 32 feature maps, the second maintains 32 maps (both followed by rectified linear unit activations), and the final convolution collapses back to a single-channel output of the same shape. We trained the CNN using the Adam optimizer at a learning rate of 0.001 to minimize the mean squared error.

D6. Comparison of Computational Cost

All analyses were conducted on a Linux server (kernel 4.18.0, x86_64 architecture) equipped with 16 CPU cores and 125.5 GB of RAM. No GPU acceleration was used. Runtime and peak memory usage are summarized in **Table D1**. In the high-adherence cohort, PMSI required 12 seconds and 0.60 MB of peak memory, compared with 1.3 seconds for LOCF, 3.2 seconds for PLI, 9 seconds for KNN, 25 seconds for CNN, and 2,328 seconds for ARIMA. Similarly, in the low-adherence cohort, PMSI required 19 seconds and 0.63 MB of peak memory, compared with 1.9 seconds for LOCF, 12 seconds for PLI, 25 seconds for KNN, 37 seconds for CNN, and 2,577 seconds for ARIMA. These results suggested that PMSI required relatively low computational cost, achieving low runtime without substantially increasing memory usage.

Table D1. Comparison of Runtime and Peak Memory Usage between PMSI and Benchmark Imputation Methods.

Method	Runtime (s)	Peak Memory Usage (MB)
High-adherence cohort (558 time series)		
PMSI	12	0.60
k-nearest neighbors (KNN)	9	0.56
ARIMA	2,328	101
Convolutional neural network (CNN)	25	0.55
Piecewise linear interpolation (PLI)	3.2	0.52
Last observation carried forward (LOCF)	1.3	0.51
Low-adherence cohort (1,957 time series)		
PMSI	19	0.63
k-nearest neighbors (KNN)	25	0.65
ARIMA	2,577	103
Convolutional neural network (CNN)	37	0.56
Piecewise linear interpolation (PLI)	12	0.53
Last observation carried forward (LOCF)	1.9	0.52

Appendix E. Ablation Analysis

Compared to the AIC-selected two-dimensional baseline, both the random-mask configuration (in which the empirically learned missingness mechanism was replaced with masks generated by random sampling) and the one-dimensional configuration (in which the multi-scale folded kernel was replaced with a single-scale kernel reflecting hourly temporal proximity only) showed an increase in both RMSE and AIC, indicating reduced accuracy and model fit (**Figure E1**). In contrast, the 3D configuration (in which the kernel was extended to capture the hourly, daily, and weekly periodicities) had a comparable RMSE, but its AIC was higher, reflecting a penalty for added model complexity that was not offset by any meaningful gain in accuracy.

The baseline kernel showed a broad spatial profile (σ_x = 1.07, σ_y = 1.88), but the random-mask kernel converged to a narrower profile (σ_x = 0.64, σ_y = 0.14) (**Figure E2**). The baseline kernel enabled it to bridge the large gaps characteristic of real-world missingness, while the random-mask kernel was consistent with optimization toward the isolated gaps produced by random masking rather than the true missingness patterns.

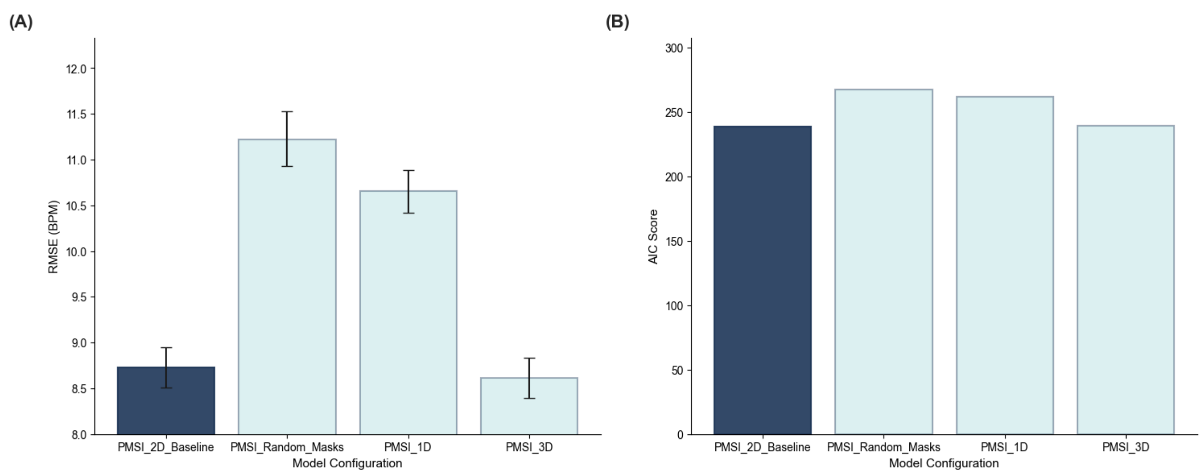

Figure E1. Ablation analysis comparing PMSI model configurations in the low-adherence cohort. The model configurations included 2D baseline (the AIC-selected two-dimensional baseline), random-mask (the empirically learned missingness mechanism replaced with masks generated by random sampling), 1D (the multi-scale folded kernel replaced with a single-scale kernel reflecting hourly temporal proximity only), and 3D (extended kernel with hourly, daily, and weekly periodicities). Results were reported using RMSE with 95% confidence intervals (**panel A**), and AIC scores (**panel B**).
Abbreviation: AIC = Akaike Information Criterion.

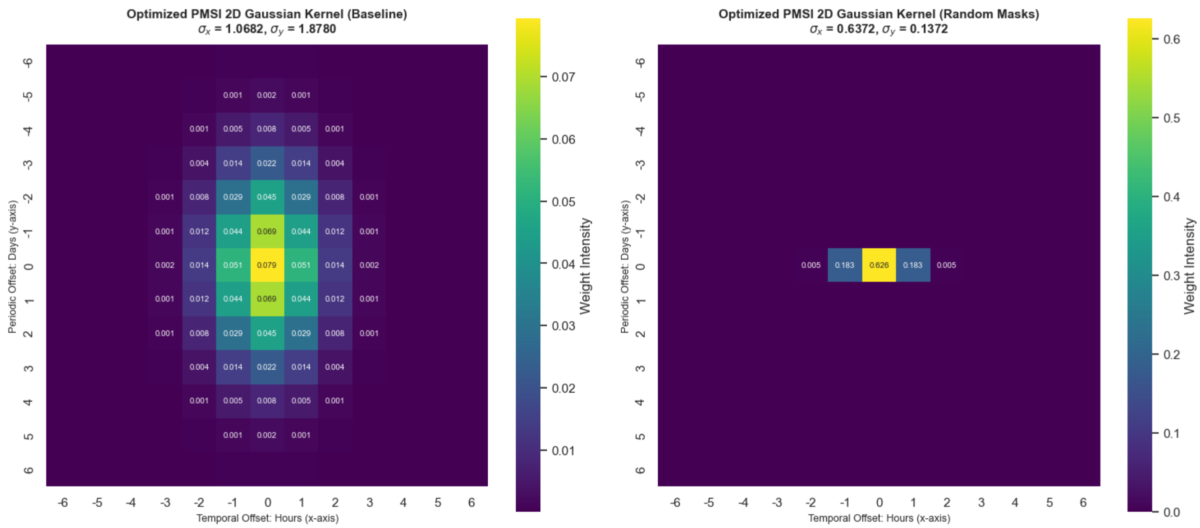

Figure E2. Optimized Gaussian kernel weights for the two-dimensional baseline (left) and random-mask (right) PMSI configurations. The random-mask PMSI configuration kept the same kernel structure as the baseline, but the empirically learned missingness mechanism was replaced with masks generated by random sampling.

Appendix F. Heart Rate Distribution

The distribution of observed heart-rate values was approximately symmetric. The mean heart rate was 70.13 bpm (with the standard deviation of 15.40), and the median was 68.77 bpm. The skewness was 0.597.

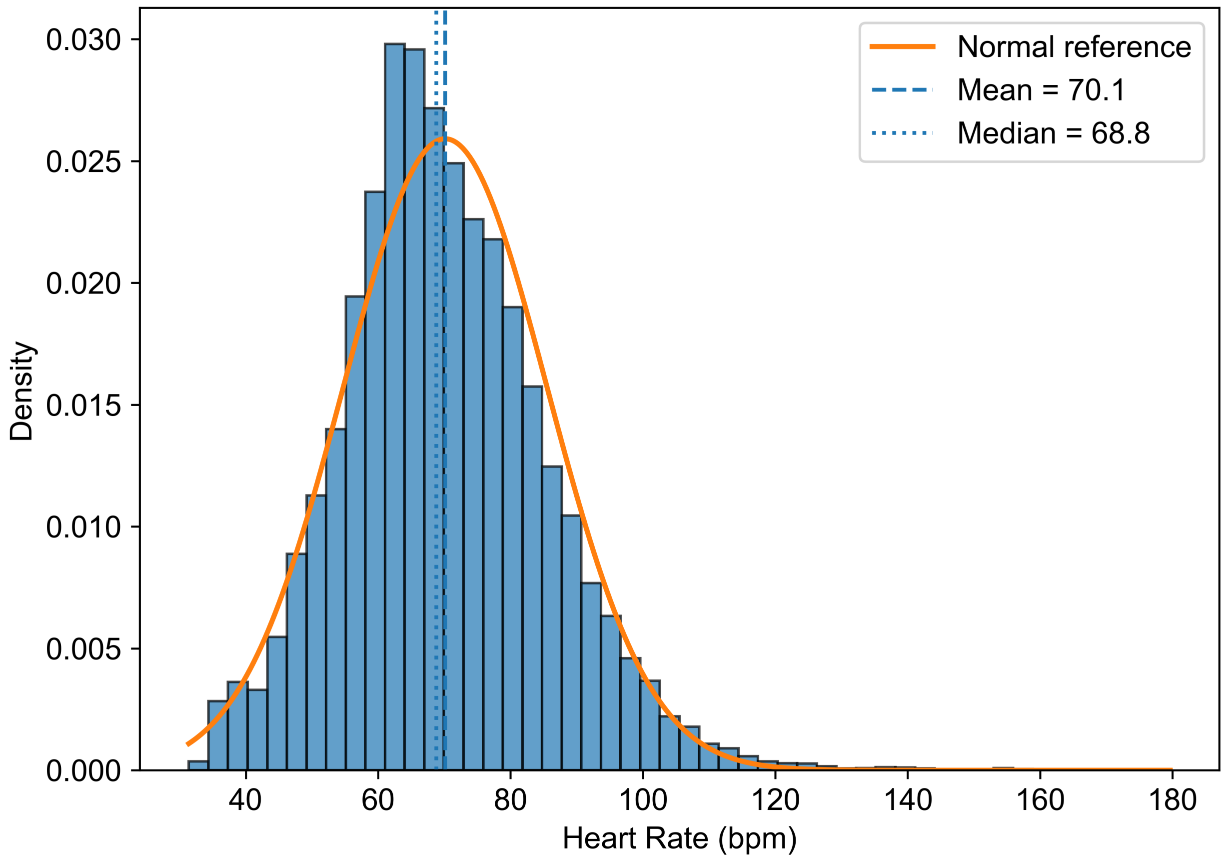

Figure F1. Distribution of heart-rate values. The empirical distribution was shown with a Gaussian reference curve; vertical lines indicated the mean and median.